



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Alternative splicing and the evolution of phenotypic novelty

Citation for published version:

Bush, SJ, Chen, L, Tovar-Corona, JM & Urrutia, AO 2017, 'Alternative splicing and the evolution of phenotypic novelty', *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1713. <https://doi.org/10.1098/rstb.2015.0474>

Digital Object Identifier (DOI):

[10.1098/rstb.2015.0474](https://doi.org/10.1098/rstb.2015.0474)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Philosophical Transactions of the Royal Society B: Biological Sciences

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



PHILOSOPHICAL TRANSACTIONS B

Alternative splicing and the evolution of phenotypic novelty

Journal:	<i>Philosophical Transactions B</i>
Manuscript ID	RSTB-2015-0474.R1
Article Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Bush, Stephen; University of Edinburgh, Roslin Institute Chen, Lu; Sichuan University West China Hospital, State Key Laboratory of Biotherapy Tovar Corona, Jaime; Wellcome Trust Sanger Institute Urrutia, Araxi; University of Bath,
Issue Code: Click here to find the code for your issue.:	DIVERSITY
Subject:	Genomics < BIOLOGY, Evolution < BIOLOGY
Keywords:	alternative splicing, functional innovation, gene duplication, comparative genomics, effective population size, genetic drift

SCHOLARONE™
Manuscripts

Alternative splicing and the evolution of phenotypic
novelty

Stephen J. Bush¹, Lu Chen², Jaime M. Tovar-Corona³, and Araxi O. Urrutia^{4,5*}

¹The Roslin Institute, University of Edinburgh, Edinburgh, EH25 9RG, UK.

²West China Second University Hospital, State Key Laboratory of Biotherapy, Sichuan University, Chengdu
610041, China.

³ Wellcome Trust Sanger Institute, Genome Campus, Hinxton, CB10 1SA, UK.

⁴Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

⁵Milner Centre for Evolution, University of Bath, Bath BA2 7AY, UK.

Keywords: Alternative splicing, functional innovation, gene duplication, comparative genomics, effective
population size, genetic drift

Running Title: Alternative splicing & phenotype novelty

Summary

Alternative splicing, a mechanism of post-transcriptional RNA processing whereby a single gene can encode multiple distinct transcripts, has been proposed to underlie morphological innovations in multicellular organisms. Genes with developmental functions are enriched for alternative splicing events, suggestive of a contribution of alternative splicing to developmental plasticity. The role of alternative splicing as a source of transcript diversification has previously been compared to that of gene duplication, with the relationship between the two extensively explored. Alternative splicing is reduced following gene duplication with the retention of duplicate copies higher for genes which were alternatively spliced prior to duplication. Furthermore, and unlike the case for overall gene number, the proportion of alternatively spliced genes has also increased in line with the evolutionary diversification of cell types, suggesting alternative splicing may contribute to the complexity of developmental programmes. Together these observations suggest a prominent role for alternative splicing as a source of functional innovation. However, it is unknown whether the proliferation of alternative splicing events indeed reflects a functional expansion of the transcriptome or instead results from weaker selection acting on larger species, which tend to have a higher number of cell types and lower population sizes.

*Author for correspondence (A.Urrutia@bath.ac.uk).

Introduction

Since life first appeared on Earth, organisms have constantly changed and diversified to populate most environments. Although far outnumbered by single-celled organisms, multicellular organisms have evolved some of the most complex structures known, the consequence of elaborate and highly regulated programmes which allow their development from a single fertilised egg. The successive refinement of these programmes, in individual lineages over many generations, underlies the observed phenotypic diversity of the extant multicellular species. All observed phenotypic adaptations are, ultimately, encoded in the genome – but identifying which changes in the DNA underlie the evolution of novel phenotypes has proven to be a difficult task. Here we review the potential relevance of alternative splicing, a mechanism of post-transcriptional RNA processing, in the regulation of various biological processes and the evolution of novel phenotypes.

Alternative splicing is a process whereby multiple functionally distinct transcripts are encoded from a single gene by the selective removal or retention of exons and/or introns from the maturing RNA [1-4]. The process is highly regulated, involving *trans*-acting splicing factors and *cis*-acting regulatory motifs (see [5] for review) and so is susceptible to hereditary and somatic mutations [6]. Alternative splicing is common in many eukaryote lineages, including metazoans [7], fungi [8] and plants [9], with deep transcriptome sequencing of the human genome showing over 95% of multi-exon genes produce at least one alternatively spliced isoform [10, 11].

Alternative splicing has been found to be involved with numerous processes in multiple taxa including the virulence of pathogenic fungi [8], the flowering time of *Arabidopsis* [12], penile erection [13] and thermogenesis [14] in mice, and the neuronal development of primates [15, 16], rats [17], worms [18] and flies [19]. Many single-gene studies have also characterised the role of alternative splicing in various cellular processes (see [5, 20] for reviews), with a recent single-cell study detecting differential alternative isoform expression throughout the cell cycle [21]. Disruption or dysregulation of alternative splicing has also been associated with pathological states (see [22, 23] for reviews) further supporting a functional contribution of this process. Aberrant alternative splicing has been linked to uncontrolled cell proliferation and chemotherapy resistance in cancer cell lines [24-28]. More broadly, genome-wide association studies in humans (which detect phenotype-predictive polymorphisms by genotyping affected and non-affected individuals) have uncovered a large number of intronic polymorphisms as significant predictors of disease susceptibility [29], many of which have been found to disrupt correct alternative splicing (e.g. [30-32]).

Some studies have associated alternative splicing with developmental plasticity and early speciation. Differential methylation patterns (which affect gene regulation) are associated with alternative splicing in bees which produce alternative castes [33]. Similarly, human body lice are thought to have emerged as an alternative morph of human head lice since the use of clothing became widespread. While there were no significant differences detected in which genes were expressed between the two morphs [34], morph-specific alternative splicing events were detected for over a thousand genes [35]. In particular, body lice specific alternative splicing events were enriched in genes associated with developmental gene ontology terms [35]. A comparable study of two mice subspecies that diverged about half a million years ago were also found to differ in their alternative splicing profiles [36].

Alternative splicing, functional innovation and the evolution of complexity in developmental programmes

Alternative splicing has been proposed to contribute to the evolution of novel phenotypes [37-40] as a small number of mutations (those in the 5' or 3' splice sites, for instance), can give rise to alternative splicing isoforms consisting of novel combinations of exons from existing genes [41, 42], providing an opportunity for the evolution of new functions. This has prompted comparisons between alternative splicing and gene duplication, widely recognised as a major source of phenotypic innovation [43, 44] (Holland, in this issue). The number of alternatively spliced isoforms per gene can decrease after a gene duplication event [45] suggesting that alternative splicing and gene duplication are to some extent equivalent mechanisms regulating transcript diversity. Several studies have reported that larger gene families (i.e. with more paralogues) have fewer alternative splicing events per gene compared to smaller families [46-48], further consistent with a reciprocal link. However, these correlations may also be explained by other factors, such as gene age (genes progressively acquire new splice variants over time, although the rate of increase is slowed by stronger purifying selection acting upon older genes) [49]. Figure 1 shows a model of the interplay between alternative splicing and gene duplication.

Alternative splicing may also promote the retention of duplicated genes by facilitating subfunctionalisation [50] where gene copies acquire reciprocal mutations so that the duplicate copies fulfil the original's gene function preventing them from becoming pseudogenes – the fate of most duplicate genes [51]. Single gene studies provide examples where duplicated copies are retained and diverge to resemble alternatively spliced isoforms produced by the ancestral single copy gene. For instance, in the vertebrate genome, which underwent at least one round of whole genome duplication early in its evolution [52, 53], there are three copies of the neuromuscular gene troponin I, each expressed in a distinct muscle type. These three genes are broadly equivalent to the alternatively spliced isoforms – also expressed in distinct muscle types – found in the urochordate *Ciona intestinalis*: a sister taxon to vertebrates, not affected by the whole genome duplication events, wherein troponin I is a single-copy gene [54]. Other examples of duplicate genes retained and evolving to resemble ancestral alternative splicing isoforms preceding gene duplication are characterised in the MITF [55], FoxP [56], and Syn-Timp families [57]. Moreover, two genome wide studies have shown that alternatively spliced genes are more likely to be found as duplicate copies in vertebrates [58] and fungi [59].

Soon after the publication of the human genome sequence – which revealed a lower than expected number of protein coding genes – alternative splicing was proposed as a candidate to explain the diversification in the number of cell types observed in some eukaryotic lineages (a higher number of cell types in a given species is assumed to reflect increased organism complexity). As gene duplication has long been associated with functional innovation, it was initially assumed that overall gene number should correlate with the number of cell types [38, 60]. Gene duplication rates, however, failed to reflect the diversification of cell types observed in several eukaryotic lineages. There is a significant correlation between the two but weaker than expected and only moderately better if the analysis is restricted to metazoans [61]. Whole genome duplication events at the base of the vertebrate lineage [52, 53], which have the highest number of cell types among eukaryotes, have comparable numbers of genes to most invertebrates (that have not undergone whole genome duplication). The poor relationship between a species' cell type diversity and its total gene number has become known as 'the G-value paradox' [60].

Although other genomic features have been shown to correlate with cell type number and may be important contributors to the evolution of complexity (i.e. [61, 62]), alternative splicing – as a mechanism allowing transcript diversification in the absence of increases in gene number – is a prime candidate to explain the G-value paradox [37, 38, 40, 63]. Comparative studies have reported marked differences in the prevalence of alternative splicing across eukaryotic lineages as well as a significant correlation between alternative splicing and the number of cell types per species [39, 61, 62, 64]. These results are in principle consistent with an adaptive role of alternative splicing in determining a genome's functional information capacity and facilitating transcript diversification in species with greater numbers of cell types. It should be noted that some of the studies exploring changes in alternative splicing across evolutionary time and its association with cell type diversification have not corrected for unequal transcript coverage between species (which distort estimates of alternative splicing [37-39, 61, 64]) leading to inconclusive results [61, 64].

Functional molecular consequences of alternative splicing

It is clear that alternative splicing makes a large contribution to the number of distinct transcripts produced in many species – but the contribution of alternative splicing to the proteome is less clear [65]. Alternative splicing can modify the properties of encoded proteins, including the set of domains it contains, its binding properties, stability, intracellular localization and enzymatic activity [66-68]. Genes with higher levels of alternative splicing tend, in general, to have a higher number of protein-protein interactions [69]; those with alternative splicing events which are tissue specific also tend to have a more central position in PPI networks [70]. Furthermore, alternatively spliced regions preferentially encode residues found at the protein surface [71], which often contain interaction sites for proteins and their binding partners [72]. In addition, a gene's alternatively spliced regions are more likely to encode disordered regions of protein (i.e. which have multiple stable configurations) [73]. These regions may also experience stronger selection for functions related to binding – flexible sensors are more adaptable than rigid ones [74, 75] – supporting an association of alternative splicing with the evolution of regulatory networks [76] (see [77] for a review of the regulatory circuits through which alternative splicing may act to regulate expression). Indeed, a comparison of the protein-protein interaction profiles of hundreds of protein isoform pairs suggests that the majority share less than half of their interactions – the principal and alternative isoforms belong to separate functional modules, suggesting functional differences between the set of splicing isoforms produced by the same gene [65]. Taken together, these observations are consistent with a role of alternative splicing in the diversification of the proteome and the modulation of protein-protein interactions.

Despite the various examples of alternatively spliced proteins having distinct functional roles, it is likely that a large proportion, or indeed the majority, of alternative splicing isoforms do not result in functional proteins [78-82]. An analysis of eight large-scale human proteomics experiments – collectively interrogating over 100 tissues, cell lines and developmental states with tandem mass spectrometry – found only 0.4% of the detected peptides are derived from alternatively spliced transcripts [83]. It is possible that the primary functional role of an alternative splicing event – if there is one at all – need not be at the protein level. For instance, alternative transcripts with 'poison cassettes' – exons with early in-frame stop codons – will never produce protein but instead down-regulate expression by diverting a proportion of pre-mRNA into the nonsense-mediated decay (NMD) pathway, a mechanism known as RUST (regulated unproductive splicing and translation) [84-86].

Selection versus genetic drift and the proliferation of alternative splicing

The increasing prevalence of alternative splicing in eukaryotes with higher number of cell types may be explained by neutral processes. In general, species with higher numbers of cell types also tend to have lower effective population sizes (the number of breeding individuals contributing to the next generation), which is thought to reduce the strength of selection relative to genetic drift [87, 88] (but see [89-91]). This is because in smaller populations there is a greater likelihood of neutral or slightly deleterious mutations – including those affecting splicing regulation – reaching fixation. Thus, it is expected that the number of ‘noisy’ transcripts (those arising by erroneous splice site recognition) should be higher in species with lower effective population sizes resulting from mutations which introduce new but non-functional alternative splicing events, or that reduce the effectiveness of accurate intron splicing, increasing stochastic noise in RNA processing [78] – physicochemical fluctuations in the cellular environment that vary spliceosome efficiency [92]. As such, a reasonable null expectation is that many of the observed alternative splicing events are selectively neutral and that their existence is as a transient state, prior to being fixed or lost due to drift.

In the absence of experimental verification of each splicing event identified, the general functional contribution of splicing events is hard to assess. Under a drift model, it would be expected that the relationship between alternative splicing and cell type number should be explained by variations in effective population size. Our previous work suggests a strong relationship between alternative splicing and cell type number irrespective of effective population size [62], however, this finding is limited by the small quantity of effective population estimates available ($n = 12$). Although compendia of effective population size data have developed substantially [93, 94], there is still minimal overlap with the set of species for which comparative alternative splicing levels have been estimated. Taking adult body weight [95], average age at sexual maturity [96], and generation time [97] as proxies of effective population size [89] as well as ‘effective information’ (a value derived from principles of information theory as ‘the minimal amount of genomic information needed to construct an organism’) [98], the correlation of alternative splicing and cell type number remains significant [62] (Table 1 and Supplementary Table 1).

More complex species with higher number of cell types will also have longer generation times. If every time a genome is replicated there is a chance of error, then ‘complex’ species (which tend to have longer generation times and smaller effective population sizes), should have fewer replications per unit time [99]. We can speculate that consequently, mutations affecting splicing regulation – such as a single nucleotide polymorphism (SNP) that creates a new splice site [100] – are less likely to arise, but that new splice sites are also less likely to be eliminated (as the strength of drift relative to selection is higher). We can also predict that in ‘complex’ species, to disrupt, or create *de novo*, a splice site is to hit a large mutational target: the binding sites for proteins involved in exon recognition (necessary for accurate intron removal) are relatively large (approx. 30bp [78]), the binding sites for RNA-binding proteins (RBPs) are, although short, numerous [101], and the introns (in which new splice sites may arise) are longer in organisms with lower effective population sizes [88].

Another way to assess the functional contribution of alternative splicing is to examine the conservation of alternatively spliced events across evolution. A base-by-base analysis of the genotype-phenotype relationship for an alternatively spliced human exon demonstrated, that nucleotides which affect splicing regulation are dispersed the entire length of the

exon and so any mutations could conceivably alter the accuracy of splicing [102]. This predicts the rapid divergence of alternative splicing events between species, except if they are selectively constrained. Conserved patterns of differential exon usage have been observed for a proportion of genes in comparisons of humans and mice [103-105], among primates [15, 106] and between grapevine cultivars [92]. Conserved alternative splicing profiles have also been identified in a lineage-, species- and tissue-specific manner among vertebrates [107, 108]. These studies suggest that in some of the species with the highest number of cell types, a significant number of the alternative splicing events are functional. However, whether the proportion of conserved alternative splicing events per species co-varies with the overall genome-wide prevalence of alternative splicing or with species complexity is unknown. In any case, increasing noise in alternative splicing regulation may, in itself, provide the substrate for a selective process and have adaptive significance to phenotypic plasticity, as it creates a 'landscape of opportunities' [109, 110].

It is likely that the observed distributions of alternative splicing events have been shaped by both adaptive and neutral processes, but further investigation is required to determine the evolutionary force of greatest significance.

Concluding remarks

The molecular drivers of development of an adult organism from a fertilised egg have been a matter of intense study over many decades. The discovery of DNA as the molecule containing the heritable information paved the way to many discoveries into the shared molecular pathways governing the development of a range of phenotypic traits. Advances in the 'post-genomic era' (characterised by its rapid accumulation of genomic, transcriptomic and proteomic profiling data) have allowed a better understanding of what molecular characteristics best explain the phenotypic differences between organisms. Perhaps more importantly, comparative studies are also helping to piece together the role of different aspects of genome evolution in the context of functional and phenotypic innovations. Alternative splicing provides an ideal mechanism for genomic novelty as it can result from a small number of mutations. Furthermore, while changes in the coding regions of genes will affect all transcripts containing this region, alternative splicing events can be expressed in highly specific conditions such as specific cell types or developmental stages. Here, we have summarised the evidence for a prominent role for alternative splicing in functional genomic evolution. Unlike gene number [61], alternative splicing is, at present, the foremost genomic predictor of increases in developmental complexity (using the number of cell types as proxy) [62]. It remains an open question as to what extent the association of alternative splicing and the number of distinct cell types is functional, or potentially a by-product of the fact that more complex organisms have lower effective population sizes and so are subject more strongly to drift.

Additional information

Data Accessibility

Primary data is included as supplementary material.

Authors' Contributions

All authors were involved in the drafting of the article and approved the final version to be published.

Competing Interests

AOU is co-guest editor of this issue and is based at the same institution as the co-editor of the issue who was also the handling editor for this manuscript.

Acknowledgements

The authors wish to thank Professor Cheryl Tickle and three anonymous reviewers for their comments and suggestions which have greatly improved the quality of this manuscript.

Funding

This work was supported by CONACyT scholarships to ACM and JMTC, a UK-China scholarship for excellence and University of Bath research studentship to LC and Royal Society Dorothy Hodgkin Fellowship DH071902, Royal Society Research Grant RG0870644 and Royal Society Grant for Fellows RG080272 to AUO.

Additional information

Supplementary table 1. Assessment of the correlation between alternative splicing and organism complexity correcting for estimates and covariates of effective population size, including generation time, average age of sexual maturity and adult body weight.

References

1. Chow L.T., Gelinas R.E., Broker T.R., Roberts R.J. 1977 An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**(1), 1-8. (doi:doi:10.1016/0092-8674(77)90180-5).

2. Berget S.M., C. M., Sharp P.A. 1977 Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**(8), 3171-3175. (doi:doi: 10.1073/pnas.74.8.3171).

3. Alt F.W., Bothwell A.L.M., Knapp M., Siden E., Mather E., Koshland M., Baltimore D. 1980 Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**(2), 293-301. (doi:doi:10.1016/0092-8674(80)90615-7).

4. Early P., Rogers J., Davis M., Calame K., Bond M., Wall R., Hood L. 1980 Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* **20**(2), 313-319. (doi:doi:10.1016/0092-8674(80)90617-0).

5. Kelemen O., Convertini P., Zhang Z., Wen Y., Shen M., Falaleeva M., Stamm S. 2013 Function of alternative splicing. *Gene* **514**(1), 1-30. (doi:10.1016/j.gene.2012.07.083).

6. Scotti M.M., Swanson M.S. 2016 RNA mis-splicing in disease. *Nat Rev Genet* **17**(1), 19-32. (doi:10.1038/nrg.2015.3).

7. Kim N., Alekseyenko A.V., Roy M., Lee C. 2007 The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Research* **35**, D93-D98. (doi:10.1093/nar/gkl884).

8. Grutzmann K., Szafranski K., Pohl M., Voigt K., Petzold A., Schuster S. 2014 Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Res* **21**(1), 27-39. (doi:10.1093/dnares/dst038).

9. Zhang C., Yang H., Yang H. 2015 Evolutionary Character of Alternative Splicing in Plants. *Bioinformatics and Biology Insights* **9**(Suppl 1), 47-52. (doi:10.4137/BBI.S33716).

10. Pan Q., Shai O., Lee L.J., Frey B.J., Blencowe B.J. 2008 Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**(12), 1413-1415. (doi:10.1038/ng.259).

11. Wang E.T., Sandberg R., Luo S.J., Khrebukova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. 2008 Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**(7221), 470-476. (doi:doi:10.1038/nature07509).

12. Macknight R., Duroux M., Laurie R., Dijkwel P., Simpson G., Dean C. 2002 Functional Significance of the Alternative Transcript Processing of the Arabidopsis Floral Promoter FCA. *The Plant Cell* **14**(4), 877-888. (doi:10.1105/tpc.010456).

13. Hurt K.J., Sezen S.F., Champion H.C., Crone J.K., Palese M.A., Huang P.L., Sawa A., Luo X., Musicki B., Snyder S.H., et al. 2006 Alternatively spliced neuronal nitric oxide synthase mediates penile erection. *Proceedings of the National Academy of Sciences of the United States of America* **103**(9), 3440-3443. (doi:10.1073/pnas.0511326103).

14. Vernia S., Edwards Y.J.K., Han M.S., Cavanagh-Kyros J., Barrett T., Kim J.K., Davis R.J. 2016 An alternative splicing program promotes adipose tissue thermogenesis. *eLife* **5**, e17672. (doi:10.7554/eLife.17672).

15. Calarco J.A., Xing Y., Caceres M., Calarco J.P., Xiao X., Pan Q., Lee C., Preuss T.M., Blencowe B.J. 2007 Global analysis of alternative splicing differences between humans and chimpanzees. *Genes & development* **21**(22), 2963-2975. (doi:10.1101/gad.1606907).

16. Lin L., Shen S., Jiang P., Sato S., Davidson B.L., Xing Y. 2010 Evolution of alternative splicing in primate brain transcriptomes. *Human molecular genetics* **19**(15), 2958-2973. (doi:10.1093/hmg/ddq201).
17. Lee H., Dean C., Isacoff E. 2010 Alternative splicing of neuroligin regulates the rate of presynaptic differentiation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30**(34), 11435-11446. (doi:10.1523/jneurosci.2946-10.2010).
18. Norris A.D., Gao S., Norris M.L., Ray D., Ramani A.K., Fraser A.G., Morris Q., Hughes T.R., Zhen M., Calarco J.A. 2014 A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol Cell* **54**(6), 946-959. (doi:10.1016/j.molcel.2014.05.004).
19. Mohr C., Hartmann B. 2014 Alternative splicing in Drosophila neuronal development. *Journal of Neurogenetics* **28**(3-4), 199-215. (doi:10.3109/01677063.2014.936437).
20. Kalsotra A., Cooper T.A. 2011 Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**(10), 715-729.
21. Welch J.D., Hu Y., Prins J.F. 2016 Robust detection of alternative splicing in a population of single cells. *Nucleic acids research* **44**(8), e73. (doi:10.1093/nar/gkv1525).
22. Cieply B., Carstens R.P. 2015 Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip Rev RNA* **6**(3), 311-326. (doi:10.1002/wrna.1276).
23. Orengo J.P., Cooper T.A. 2007 Alternative splicing in disease. *Adv Exp Med Biol* **623**, 212-223.
24. Venables J.P. 2004 Aberrant and alternative splicing in cancer. *Cancer research* **64**(21), 7647-7654.
25. Venables J.P., Klinck R., Koh C., Gervais-Bird J., Bramard A., Inkel L., Durand M., Couture S., Froehlich U., Lapointe E. 2009 Cancer-associated regulation of alternative splicing. *Nature structural & molecular biology* **16**(6), 670-676.
26. Xu Q., Lee C. 2003 Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic acids research* **31**(19), 5635-5643.
27. Chen L., Tovar-Corona J.M., Urrutia A.O. 2011 Increased levels of noisy splicing in cancers, but not for oncogene-derived transcripts. *Human molecular genetics* **20**(22), 4422-4429. (doi:10.1093/hmg/ddr370).
28. Brinkman B.M.N. 2004 Splice variants as cancer biomarkers. *Clinical biochemistry* **37**(7), 584-594.
29. Xiong H.Y., Alipanahi B., Lee L.J., Bretschneider H., Merico D., Yuen R.K., Hua Y., Gueroussov S., Najafabadi H.S., Hughes T.R. 2015 The human splicing code reveals new insights into the genetic determinants of disease. *Science (New York, NY)* **347**(6218), 1254806.
30. Prokunina-Olsson L., Welch C., Hansson O., Adhikari N., Scott L.J., Usher N., Tong M., Sprau A., Swift A., Bonnycastle L.L. 2009 Tissue-specific alternative splicing of TCF7L2. *Human molecular genetics* **18**(20), 3795-3804.
31. Mondal A.K., Das S.K., Baldini G., Chu W.S., Sharma N.K., Hackney O.G., Zhao J., Grant S.F., Elbein S.C. 2010 Genotype and tissue-specific effects on alternative splicing of the transcription factor 7-like 2 gene in humans. *The Journal of Clinical Endocrinology & Metabolism* **95**(3), 1450-1457.
32. Burkhardt R., Kenny E.E., Lowe J.K., Birkeland A., Josowitz R., Noel M., Salit J., Maller J.B., Pe'er I., Daly M.J. 2008 Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arteriosclerosis, thrombosis, and vascular biology* **28**(11), 2078-2084.
33. Lyko F., Foret S., Kucharski R., Wolf S., Falckenhayn C., Maleszka R. 2010 The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* **8**(11), e1000506.
34. Olds B., Coates B., Steele L., Sun W., Agunbiade T., Yoon K., Strycharz J., Lee S., Paige K., Clark J. 2012 Comparison of the transcriptional profiles of head and body lice. *Insect molecular biology* **21**(2), 257-268.
35. Tovar-Corona J.M., Castillo-Morales A., Chen L., Olds B.P., Clark J.M., Reynolds S.E., Pittendrigh B.R., Feil E.J., Urrutia A.O. 2015 Alternative splice in alternative lice. *Molecular biology and evolution* **32**(10), 2749-2759.
36. Harr B., Turner L.M. 2010 Genome-wide analysis of alternative splicing evolution among Mus subspecies. *Molecular ecology* **19**(s1), 228-239.
37. Chen L., Tovar-Corona J.M., Urrutia A.O. 2012 Alternative Splicing: A Potential Source of Functional Innovation in the Eukaryotic Genome. *Int Journal Evol Biol* **2012**, 10. (doi:10.1155/2012/596274).
38. Nilsen T.W., Graveley B.R. 2010 Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**(7280), 457-463. (doi:10.1038/nature08909).
39. Kim E., Magen A., Ast G. 2007 Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* **35**(1), 125-131. (doi:10.1093/nar/gkl924).
40. Graveley B.R. 2001 Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**(2), 100-107. (doi:[http://dx.doi.org/10.1016/S0168-9525\(00\)02176-4](http://dx.doi.org/10.1016/S0168-9525(00)02176-4)).
41. Krawczak M., Reiss J., Cooper D.N. 1992 The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Human genetics* **90**(1-2), 41-54.
42. Ward A.J., Cooper T.A. 2010 The Pathobiology of Splicing. *The Journal of pathology* **220**(2), 152-163. (doi:10.1002/path.2649).
43. Lynch M., Conery J.S. 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290**(5494), 1151-1155.
44. Krylov D.M., Wolf Y.I., Rogozin I.B., Koonin E.V. 2003 Gene Loss, Protein Sequence Divergence, Gene Dispensability, Expression Level, and Interactivity Are Correlated in Eukaryotic Evolution. *Genome Research* **13**(10), 2229-2235. (doi:10.1101/gr.1589103).
45. Su Z., Wang J., Yu J., Huang X., Gu X. 2006 Evolution of alternative splicing after gene duplication. *Genome research* **16**(2), 182-189. (doi:10.1101/gr.4197006).

46. Kopelman N.M., Lancet D., Yanai I. 2005 Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nature genetics* **37**(6), 588-589. (doi:http://www.nature.com/ng/journal/v37/n6/supinfo/ng1575_S1.html).
47. Su Z., Wang J., Yu J., Huang X., Gu X. 2006 Evolution of alternative splicing after gene duplication. *Genome Research* **16**(2), 182-189. (doi:10.1101/gr.4197006).
48. Talavera D., Vogel C., Orozco M., Teichmann S.A., de la Cruz X. 2007 The (In)dependence of Alternative Splicing and Gene Duplication. *PLoS computational biology* **3**(3), e33. (doi:10.1371/journal.pcbi.0030033).
49. Roux J., Robinson-Rechavi M. 2011 Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication. *Genome Research* **21**(3), 357-363. (doi:10.1101/gr.113803.110).
50. Su Z., Gu X. 2012 Revisit on the evolutionary relationship between alternative splicing and gene duplication. *Gene* **504**(1), 102-106.
51. Lynch M., Force A. 2000 The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**(1), 459-473.
52. Dehal P., Boore J.L. 2005 Two rounds of whole genome duplication in the ancestral vertebrate. *Plos Biology* **3**(10), 1700-1708.
53. McLysaght A., Hokamp K., Wolfe K.H. 2002 Extensive genomic duplication during early chordate evolution. *Nature genetics* **31**(2), 200-204.
54. MacLean D.W., Meedel T.H., Hastings K.E.M. 1997 Tissue-specific alternative splicing of ascidian troponin I isoforms - Redesign of a protein isoform-generating mechanism during chordate evolution. *Journal of Biological Chemistry* **272**(51), 32115-32120.
55. Altschmied J., Delfgaauw J., Wilde B., Duschl J., Bouneau L., Volff J.N., Scharl M. 2002 Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics* **161**(1), 259-267.
56. Santos M.E., Athanasiadis A., Leitão A.B., DuPasquier L., Sucena É. 2011 Alternative Splicing and Gene Duplication in the Evolution of the FoxP Gene Subfamily. *Molecular Biology and Evolution* **28**(1), 237-247. (doi:10.1093/molbev/msq182).
57. Yu W.P., Brenner S., Venkatesh B. 2003 Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. *Trends Genet* **19**(4), 180-183. (doi:10.1016/s0168-9525(03)00048-9).
58. Lambert M.J., Cochran W.O., Wilde B.M., Olsen K.G., Cooper C.D. 2015 Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. *Genome Res* **25**(5), 624-632. (doi:10.1101/gr.184473.114).
59. Marshall A.N., Montealegre M.C., Jimenez-Lopez C., Lorenz M.C., van Hoof A. 2013 Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet* **9**(3), e1003376. (doi:10.1371/journal.pgen.1003376).
60. Hahn M.W., Wray G.A. 2002 The g-value paradox. *Evolution and Development* **4**(2), 73-75.
61. Schad E., Tompa P., Hegyi H. 2011 The relationship between proteome size, structural disorder and organism complexity. *Genome Biol* **12**(12), R120.
62. Chen L., Bush S.J., Tovar-Corona J.M., Castillo-Morales A., Urrutia A.O. 2014 Correcting for Differential Transcript Coverage Reveals a Strong Relationship between Alternative Splicing and Organism Complexity. *Molecular Biology and Evolution* **31**(6), 1402-1413. (doi:10.1093/molbev/msu083).
63. Xing Y., Lee C. 2007 Relating alternative splicing to proteome complexity and genome evolution. *Advances in experimental medicine and biology* **623**, 36-49.
64. Warnefors M., Eyre-Walker A. 2011 The accumulation of gene regulation through time. *Genome biology and evolution* **3**, 667-673.
65. Yang X., Coulombe-Huntington J., Kang S., Sheynkman G.M., Hao T., Richardson A., Sun S., Yang F., Shen Y.A., Murray R.R., et al. 2016 Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**(4), 805-817. (doi:10.1016/j.cell.2016.01.029).
66. Resch A., Xing Y., Modrek B., Gorlick M., Riley R., Lee C. 2004 Assessing the impact of alternative splicing on domain interactions in the human proteome. *Journal of proteome research* **3**(1), 76-83. (doi:doi: 10.1021/pr034064v).
67. Xing Y., Xu Q., Lee C. 2003 Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS letters* **555**(3), 572-578. (doi:doi: 10.1016/S0014-5793(03)01354-1).
68. Stamm S., Ben-Ari S., Rafalska I., Tang Y., Zhang Z., Toiber D., Thanaraj T.A., Soreq H. 2005 Function of alternative splicing. *Gene* **344**, 1-20. (doi:10.1016/j.gene.2004.10.022).
69. Neverov A.D., Artamonova, II, Nurtdinov R.N., Frishman D., Gelfand M.S., Mironov A.A. 2005 Alternative splicing and protein function. *BMC bioinformatics* **6**, 266. (doi:10.1186/1471-2105-6-266).
70. Ellis J.D., Barrios-Rodiles M., Colak R., Irimia M., Kim T., Calarco J.A., Wang X., Pan Q., O'Hanlon D., Kim P.M., et al. 2012 Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell* **46**(6), 884-892. (doi:10.1016/j.molcel.2012.05.037).
71. Wang P., Yan B., Guo J.T., Hicks C., Xu Y. 2005 Structural genomics analysis of alternative splicing and application to isoform structure modeling. *Proceedings of the National Academy of Sciences of the United States of America* **102**(52), 18920-18925. (doi:10.1073/pnas.0506770102).
72. Yura K., Shionyu M., Hagino K., Hijikata A., Hirashima Y., Nakahara T., Eguchi T., Shinoda K., Yamaguchi A., Takahashi K., et al. 2006 Alternative splicing in human transcriptome: functional and structural influence on proteins. *Gene* **380**(2), 63-71. (doi:10.1016/j.gene.2006.05.015).
73. Romero P.R., Zaidi S., Fang Y.Y., Uversky V.N., Radivojac P., Oldfield C.J., Cortese M.S., Sickmeier M., LeGall T., Obradovic Z., et al. 2006 Alternative splicing in concert with protein intrinsic disorder enables increased

- functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America* **103**(22), 8390-8395. (doi:10.1073/pnas.0507916103).
74. Dunker A.K., Brown C.J., Lawson J.D., Iakoucheva L.M., Obradovic Z. 2002 Intrinsic disorder and protein function. *Biochemistry* **41**(21), 6573-6582.
75. Romero P., Obradovic Z., Dunker A.K. 2004 Natively disordered proteins: functions and predictions. *Appl Bioinformatics* **3**(2-3), 105-113.
76. Blencowe B.J. 2006 Alternative Splicing: New Insights from Global Analyses. *Cell* **126**(1), 37-47. (doi:<http://dx.doi.org/10.1016/j.cell.2006.06.023>).
77. Jangi M., Sharp P.A. 2014 Building robust transcriptomes with master splicing factors. *Cell* **159**(3), 487-498. (doi:10.1016/j.cell.2014.09.054).
78. Pickrell J.K., Pai A.A., Gilad Y., Pritchard J.K. 2010 Noisy splicing drives mRNA isoform diversity in human cells. *PLoS genetics* **6**(12), e1001236. (doi:10.1371/journal.pgen.1001236).
79. Leoni G., Le Pera L., Ferre F., Raimondo D., Tramontano A. 2011 Coding potential of the products of alternative splicing in human. *Genome Biol* **12**(1), R9.
80. Sorek R., Shamir R., Ast G. 2004 How prevalent is functional alternative splicing in the human genome? *Trends Genet* **20**(2), 68-71. (doi:10.1016/j.tig.2003.12.004).
81. Melamud E., Moulton J. 2009 Stochastic noise in splicing machinery. *Nucleic Acids Res* **37**(14), 4873-4886. (doi:10.1093/nar/gkp471).
82. Hon C.C., Weber C., Sismeiro O., Proux C., Koutero M., Deloger M., Das S., Agrahari M., Dillies M.A., Jagla B., et al. 2013 Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. *Nucleic acids research* **41**(3), 1936-1952. (doi:10.1093/nar/gks1271).
83. Tress M.L., Abascal F., Valencia A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends in Biochemical Sciences*. (doi:10.1016/j.tibs.2016.08.008).
84. Lareau L.F., Brenner S.E. 2015 Regulation of Splicing Factors by Alternative Splicing and NMD Is Conserved between Kingdoms Yet Evolutionarily Flexible. *Molecular Biology and Evolution* **32**(4), 1072-1079. (doi:10.1093/molbev/msv002).
85. Lareau L.F., Brooks A.N., Soergel D.A., Meng Q., Brenner S.E. 2007 The coupling of alternative splicing and nonsense-mediated mRNA decay. *Adv Exp Med Biol* **623**, 190-211.
86. Lewis B.P., Green R.E., Brenner S.E. 2003 Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America* **100**(1), 189-192. (doi:10.1073/pnas.0136770100).
87. Lynch M. 2007 The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences* **104**(suppl 1), 8597-8604. (doi:10.1073/pnas.0702207104).
88. Lynch M., Conery J.S. 2003 The origins of genome complexity. *Science (New York, NY)* **302**(5649), 1401-1404. (doi:10.1126/science.1089370).
89. Wu X., Hurst L.D. 2015 Why Selection Might Be Stronger When Populations Are Small: Intron Size and Density Predict within and between-Species Usage of Exonic Splice Associated cis-Motifs. *Molecular Biology and Evolution* **32**(7), 1847-1861. (doi:10.1093/molbev/msv069).
90. Whitney K.D., Garland Jr T. 2010 Did genetic drift drive increases in genome complexity? *PLoS genetics* **6**(8), e1001080.
91. Whitney K.D., Boussau B., Baack E.J., Garland Jr T. 2011 Drift and genome complexity revisited. *PLoS genetics* **7**(6), e1002092.
92. Potenza E., Racchi M.L., Sterck L., Coller E., Asquini E., Tosatto S.C.E., Velasco R., Van de Peer Y., Cestaro A. 2015 Exploration of alternative splicing events in ten different grapevine cultivars. *BMC Genomics* **16**(1), 706. (doi:10.1186/s12864-015-1922-5).
93. Palstra F.P., Ruzzante D.E. 2008 Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular ecology* **17**(15), 3428-3447.
94. Palstra F.P., Fraser D.J. 2012 Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and evolution* **2**(9), 2357-2365. (doi:10.1002/ece3.329).
95. Pelletier F., Clutton-Brock T., Pemberton J., Tuljapourkar S., Coulson T. 2007 The evolutionary demography of ecological change: linking trait variation and population growth. *Science (New York, NY)* **315**(5818), 1571-1574.
96. Lee A.M., Engen S., Sæther B.-E. 2011 The influence of persistent individual differences and age at maturity on effective population size. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20110283.
97. Chao L., Carr D.E. 1993 The molecular clock and the relationship between population-size and generation time. *Evolution* **47**(2), 688-690. (doi:10.2307/2410082).
98. Jiang Y., Xu C. 2010 The calculation of information and organismal complexity. *Biology direct* **5**, 59-59. (doi:10.1186/1745-6150-5-59).
99. Thomas J.A., Welch J.J., Lanfear R., Bromham L. 2010 A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. *Molecular Biology and Evolution* **27**(5), 1173-1180. (doi:10.1093/molbev/msq009).
100. Lynch M. 2010 Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences* **107**(3), 961-968. (doi:10.1073/pnas.0912629107).
101. Shulman-Peleg A., Nussinov R., Wolfson H.J. 2009 RsiteDB: a database of protein binding pockets that interact with RNA nucleotide bases. *Nucleic acids research* **37**(Database issue), D369-373. (doi:10.1093/nar/gkn759).
102. Julien P., Minana B., Baeza-Centurion P., Valcarcel J., Lehner B. 2016 The complete local genotype-phenotype landscape for the alternative splicing of a human exon. *Nat Commun* **7**, 11558. (doi:10.1038/ncomms11558).

103. Thanaraj T.A., Clark F., Muilu J. 2003 Conservation of human alternative splice events in mouse. *Nucleic Acids Research* **31**(10), 2544-2552. (doi:10.1093/nar/gkg355).

104. Yeo G.W., Van Nostrand E., Holste D., Poggio T., Burge C.B. 2005 Identification and analysis of alternative splicing events conserved in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America* **102**(8), 2850-2855. (doi:10.1073/pnas.0409742102).

105. Nurtdinov R.N., Artamonova I.I., Mironov A.A., Gelfand M.S. 2003 Low conservation of alternative splicing patterns in the human and mouse genomes. *Human molecular genetics* **12**(11), 1313-1320. (doi:10.1093/hmg/ddg137).

106. Reyes A., Anders S., Weatheritt R.J., Gibson T.J., Steinmetz L.M., Huber W. 2013 Drift and conservation of differential exon usage across tissues in primate species. *Proceedings of the National Academy of Sciences*. (doi:10.1073/pnas.1307202110).

107. Barbosa-Morais N.L., Irimia M., Pan Q., Xiong H.Y., Gueroussov S., Lee L.J., Slobodeniuc V., Kutter C., Watt S., Colak R., et al. 2012 The evolutionary landscape of alternative splicing in vertebrate species. *Science (New York, NY)* **338**(6114), 1587-1593. (doi:10.1126/science.1230612).

108. Merkin J., Russell C., Chen P., Burge C.B. 2012 Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science (New York, NY)* **338**(6114), 1593-1599. (doi:10.1126/science.1228186).

109. Skandalis A., Frampton M., Seger J., Richards M.H. 2010 The adaptive significance of unproductive alternative splicing in primates. *RNA* **16**(10), 2014-2022. (doi:10.1261/rna.2127910).

110. Kianianmomeni A., Ong C.S., Ratsch G., Hallmann A. 2014 Genome-wide analysis of alternative splicing in *Volvox carteri*. *BMC Genomics* **15**, 1117. (doi:10.1186/1471-2164-15-1117).

Tables

Table 1. Spearman Rho correlation coefficients between the average number of alternative splicing events per gene (ASL) and the proportion of genes alternatively spliced (ASP) with two measures of complexity (effective information and cell type number) correcting by various estimates and proxies of effective population size (Ne).

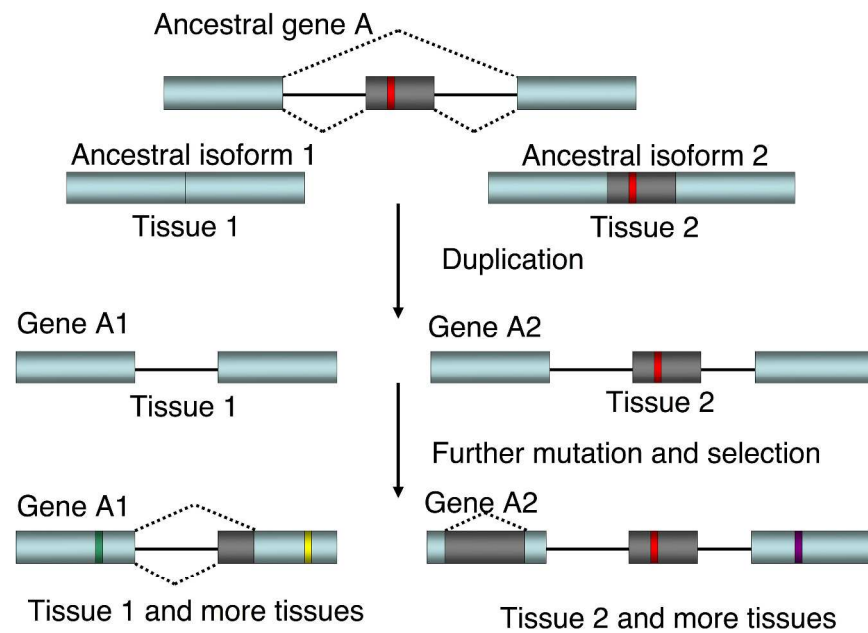
	ASL vs. Ne	ASL vs. complexity correcting by Ne		ASP vs. Ne	ASP vs. complexity correcting by Ne	
		Effective information	Cell type number		Effective information	Cell type number
Ne. μ , calculated using π/π	0.22	0.90	0.88	0.27	0.86	0.93
Ne. μ , calculated using S	0.28	0.88	0.88	0.34	0.83	0.93
Ne. μ , calculating using η/η	0.26	0.90	0.88	0.32	0.86	0.93
Adult body weight	0.64	0.21	0.93	0.39	0.55	0.89
Average age at sexual maturity	0.93	0.56	0.97	0.88	0.79	0.76
Generation time	0.70	0.86	0.78	0.73	0.85	0.82

Notes. Significant correlation coefficients significant at $p < 0.05$ are marked in bold. See supplementary table 1 for underlying data and sources.

Figure caption

Figure 1. Model of the role of alternative splicing as a source of functional genomic innovation and retention of duplicate genes by facilitating sub-functionalisation. Novel alternative splicing events allow single-copy gene A to produce two protein isoforms, each optimised for expression in a specific tissue. After a duplication event the alternatively spliced exon is maintained in gene A2 but lost in A1. Divergence in the promoter areas of the duplicated genes lead to changes in gene expression profiles over time so that each gene is now only expressed in one tissue, mirroring the expression profile of the two isoforms of the original ancestral single-copy gene.

For Review Only



2449x1715mm (72 x 72 DPI)